UNITED STATES DEPARTMENT OF THE INTERIOR

GEOLOGICAL SURVEY

Developing an Oil Generation Model for Resource Assessment of the Bakken Formation, U.S. Portion of the Williston Basin

by

Katherine B. Krystinik[1]
and
Ronald R. Charpentier[1]

Open-File Report 84-694

This report is preliminary and has not been reviewed for conformity with U.S. Geological Survey editorial standards and stratigraphic nomenclature.

[1] Denver, CO

CONTENTS

	Page
Abstract	1
Introduction	1
Data	3
Analysis of variability within the data set	5
Principal components	5
TOC vs density	6
Cluster and discriminant analysis	9
Separation of maturity effects from deposition	14
Conclusions	20
Acknowledgments	20
References	23

ILLUSTRATIONS

		Page
1.	Locations of wells where well-log data were obtained, and approximate limits of upper, middle, and lower members of Bakken Formation (from Hester and Schmoker, 1983)	2
2.	Locations of wells where geochemical data were obtained. Data from Webster (1982) and Price	4
3.	First principal component scores for upper member	7
4.	Total organic carbon versus density	10
5.	Values of first discriminant function versus values of second discriminant function showing separation of the four groups	12
6.	Scores on first discriminant function for upper member	13
7.	Density versus formation temperature and regression of density on formation temperature for upper member. (R^2=0.51)	15
8.	Density versus formation temperature and regression of density on formation temperature for lower member. (R^2=0.32)	16
9.	Residuals from regression of density on formation temperature for upper member	17
10.	Alternate models--linear and curvilinear--for the relationship between total organic carbon and temperature	18
11.	Present-day temperature (in degrees Fahrenheit) of Bakken Formation (from Hester and Schmoker, 1983)	19
12.	Interpretation of original depositional pattern of total organic carbon for upper member	21
13.	Interpretation of present-day distribution pattern of total organic carbon in upper member	22

TABLES

Page

1. Principal components coefficients--upper member. The first principal component explains 51 percent of the variability and the second principal component explains 26 percent of the variability..... 6
2. Correlation coefficients for regressions using total organic carbon and density..... 9
3. Discriminant function coefficients--upper member..... 11

DEVELOPING AN OIL GENERATION MODEL FOR RESOURCE ASSESSMENT OF THE BAKKEN FORMATION, U.S. PORTION OF THE WILLISTON BASIN

by

Katherine B. Krystinik
and
Ronald R. Charpentier

ABSTRACT

A study of the Bakken Formation, the proposed source rock for much of the hydrocarbons generated in the Williston basin, was done using well-log data. Principal components analysis, cluster analysis, and discriminant analysis were used on bulk density, neutron porosity, and resistivity logs, and formation temperatures. These analyses indicate that the present-day distribution of organic matter controls much of the variability in the log values. The pattern of present-day total organic carbon (TOC) is high in the central part of the basin near northeastern Montana and along the east edge of the basin. Low values of TOC occur in the area of the Nesson anticline and along the southwest edge of the basin. Using the regression of density on temperature and the analysis of residuals from this regression, it is possible to separate maturity effects from those of original deposition. These analyses reveal that original concentrations of organic matter were low near the shoreline and increased offshore to a high in northeast Montana. The pre-maturation and present-day TOC distributions derived using statistical analyses and well-log data can easily be explained by the depositional pattern and thermal history that would be expected in this basin, and by geochemical analyses.

INTRODUCTION

The Devonian-Mississippian Bakken Formation has been proposed as one of the major petroleum source rocks in the Williston basin (Williams, 1974; Dow, 1974). A classic study on oil generation from the Bakken was done by Meissner (1978). The Bakken consists of three informal members: upper and lower organic-rich shale members and a silty middle member. The Bakken exists only in the subsurface, but extends over most of North Dakota, the eastern part of Montana, and into Canada. The formation, at its thickest, is almost 140 feet thick. The lower shale member is the least extensive (fig. 1) and is up to 45 feet thick. The middle member, more extensive than the lower (fig. 1), has a maximum thickness of more than 70 feet. The upper shale member is the most extensive (fig. 1) but has a maximum thickness of only about 20 feet.

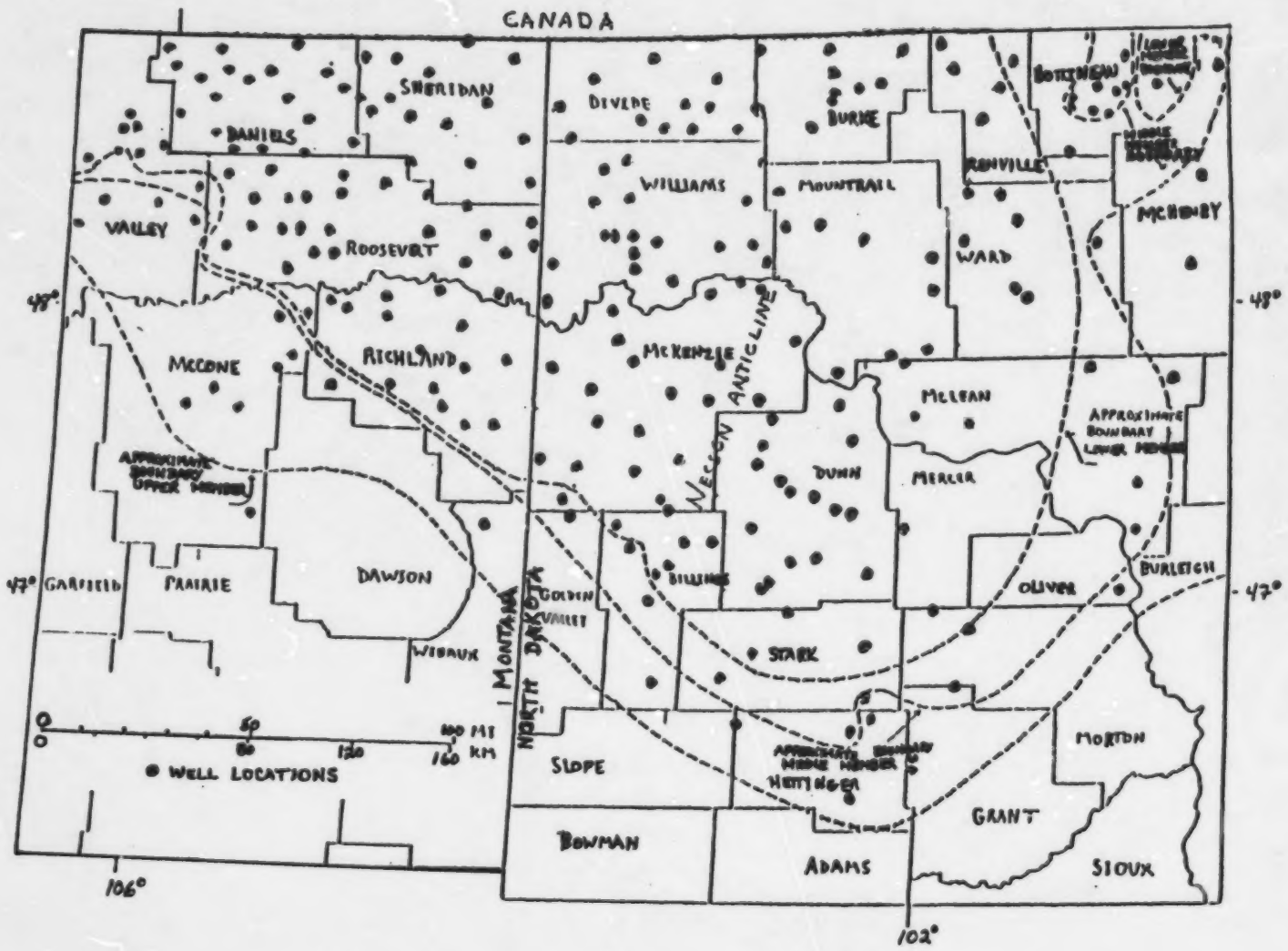

CANADA
SHERIDAN
DANIELS
VALLEY
ROOSEVELT
DIVIDE
BURKE
WILLIAMS
MOUNTRAIL
RENVILLE
BOTTINEAU
MCHENRY
WARD
48°
MCCONE
RICHLAND
MCKENZIE
ANTICLINE
WESTON
MCLEAN
DUNN
MERCER
APPROXIMATE BOUNDARY LOWER MEMBER
APPROXIMATE BOUNDARY UPPER MEMBER
47°
GARFIELD
PRAIRIE
DAWSON
WIBAUX
BILLINGS
GOLDEN VALLEY
OLIVER
BURLEIGH
47°
STARK
MONTANA
NORTH DAKOTA
0 60 120 160 KM
0 60 120 200 MI
WELL LOCATIONS
SLOPE
HETTINGER
MORTON
GRANT
BOWMAN
ADAMS
SIOUX
106°
102°

Figure 1.

One method of assessing the hydrocarbon resources of a basin is to begin with a quantitative estimate of the amount of hydrocarbon generated, and then to estimate the proportion of that hydrocarbon that was trapped. The estimation of generated hydrocarbon amount is dependent on a quantitative model of the history of the organic matter. This study provides the basis for such a model by deriving the distributions of organic matter both before and after maturation.

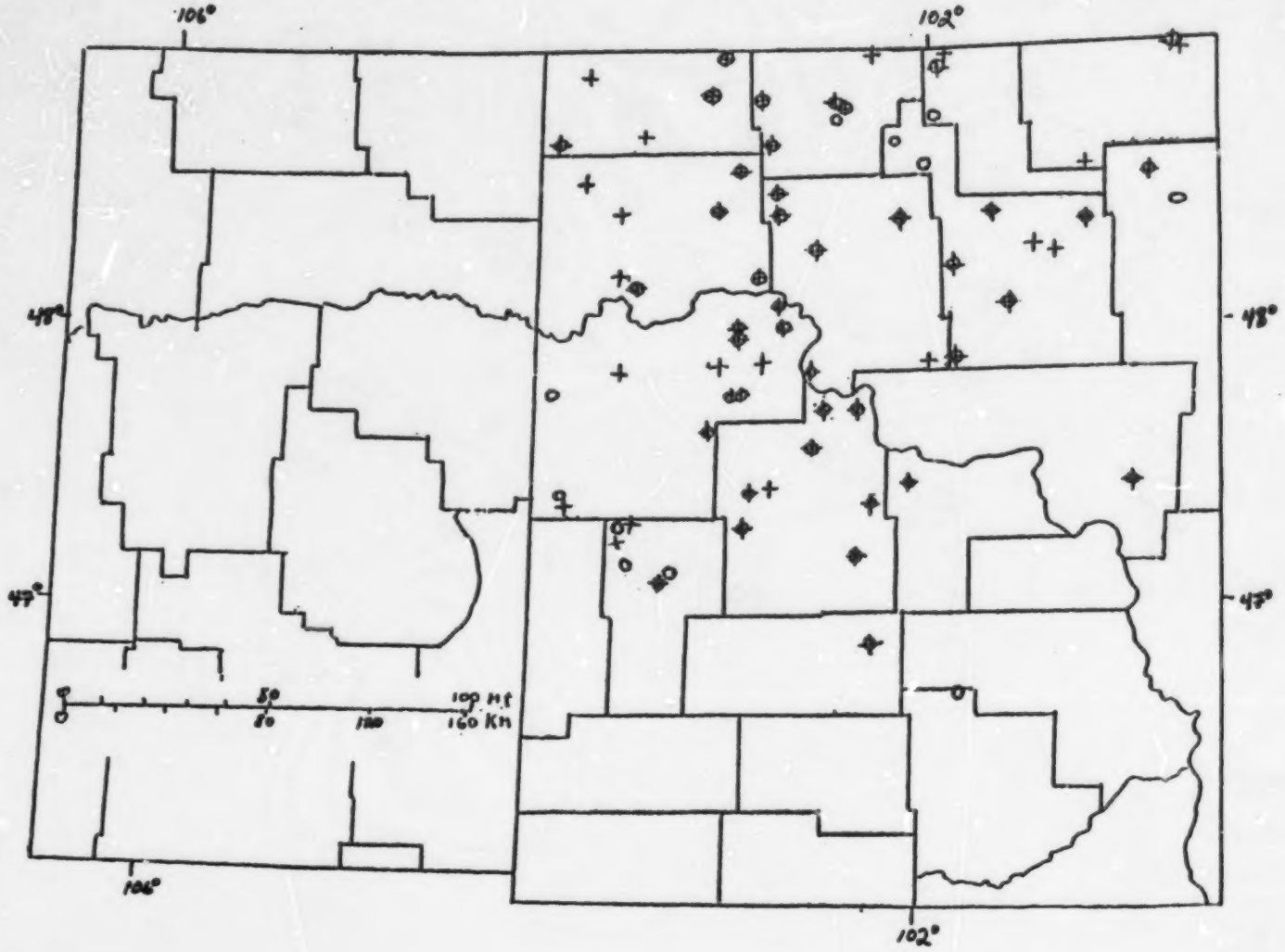
Well-log data was used in this study to estimate the pre- and post-maturation patterns of organic matter. One aim of this study was to determine if well-log data could be successfully used to make such estimates. This is of practical interest because well-log data are more readily available than geochemical data.

DATA

Schmoker and Hester (1983) visually estimated mean-log values for each member of the Bakken Formation in 254 wells (fig. 1). The log suite included, where available, bulk density log, neutron porosity log, gamma-ray intensity log, resistivity log, and sonic log. Wells at the edge of the lateral extent of each member often had unusable log values because of the thinness of the beds, but these wells were useful in determining the lateral extent of each member. There were about 170 wells with a usable suite of log values in the upper member and 105 in the lower member. The data set contained information on well identification and location; tops, bases, and thicknesses of members; and bottom-hole and formation temperatures. It also included estimates of total organic carbon (TOC) derived by Schmoker and Hester using density log values.

Geochemical data came from two sources. Price provided unpublished pyrolysis analyses of 171 samples from 52 wells and 40 soxhlet analyses from 30 wells (fig. 2). This data was later published in Price and others (1984). Webster (1982) provided various analyses on 83 samples from 61 wells (fig. 2). Webster's analyses included 57 by thin-layer chromatography, 55 by pyrolysis, 42 by visual kerogen typing, and 8 by vitrinite reflectance. Most of Webster's pyrolysis data were the results of the same analyses as those provided by Price.

Crossplots of some of the well-log data were constructed to aid in detection of suspect data. From these plots, 10 wells in the lower member and 2 in the upper member were shown to have suspect values. Each of these wells had shale thickness of less than 10 feet and most were at the edge of the member. These points were not used in later analyses.



O PRICE

+ WEBSTER (1982)

FIGURE 2.

ANALYSIS OF VARIABILITY WITHIN THE DATA SET

The first part of this study consisted of an analysis of the data to determine the causes of variability within the data set and, if possible, to relate parts of the variability to factors important to a model of oil generation. To this end, principal components analysis, regression analysis, cluster analysis, and discriminant analysis were used.

Modified log variables were used in the principal components, cluster, and discriminant analyses. These modified variables were the result of regressing each original variable on depth. The component of each original variable linearly related to depth was then removed from the original variable. Each depth-removed variable was standardized by subtracting its mean and dividing by its standard deviation. Depth was removed because, in a preliminary principal components analysis, depth dominated all other depositional or maturity effects. The variables were standardized to ensure that no variable would dominate others simply because of the size of the readings.

Principal Components

Principal components analysis was used to determine groups of well-log variables that represent the variability in the data. Principal components take the form of a linear combination of the variables. The first principal component indicates the greatest source of variability in the data. Once this variability is removed from the data, the greatest source of variabiity remaining in the data is the second principal component. (Readers desiring a more theoretical discussion of the technique may refer to any multivariate statistical text, for instance Green, 1978).

Principal components results on the Bakken log variables yield two components that explain approximately 77 percent of the variance for the upper member (table 1). Results for the lower member are similar. Variables used in the analysis were a modified density, a modified neutron porosity, a modified formation temperature, and a modified resistivity. The first principal component is a linear combination of density and formation temperature, each with a positive coefficient, and neutron porosity with a negative coefficient. The first principal component explains 51 percent of the variability in the log variables. The second principal component is predominantly resistivity.

Table 1.--Principal components coefficients--upper member. The first principal component explains 51 percent of the variability and the second principal component explains 26 percent of the variability.

	First principal component	Second principal component
density	0.88	-0.08
neutron porosity	-0.86	0.22
resistivity	0.12	0.97
temperature	0.73	0.19

The first principal component scores for the upper member show a distinct pattern (fig. 3). Low values of the principal component scores occur on the east edge and in the central part of the basin near northeastern Montana. High principal component scores occur along the southwest edge and in the area of the Nesson anticline. This pattern is similar to that of TOC as shown by Price and others (1984) and Webster (1982) from lab analyses of TOC and as predicted by Schmoker and Hester (1983) using density log values. The pattern of first principal component scores for the lower member is similar but not as clear. The first principal component is high where the TOC map is low and vice versa due to the signs of the coefficients in the first principal component.

The strong correlation between the first principal component and TOC is reasonable. Organic matter, being less dense than other matrix materials, should be negatively correlated with bulk density. This relation is discussed further in the next section. Higher temperatures could have led to a reduction of organic content by generation and expulsion of hydrocarbons. Also, since neutron porosity logs reflect the frequency of hydrogen atoms in the rock, neutron porosity readings should, in part, reflect the amount of hydrogen-rich organic matter in the rock.

TOC versus Density

Quantitative support exists for the relation between TOC and density. Schmoker and Hester (1983) derived a linear equation relating weight-percent organic carbon and the inverse of bulk density. This equation was based on (1) the relation between bulk density and densities of the individual rock components (rock matrix, interstitial pores, pyrite, and organic matter), (2) the relation between weight-percent and volume-percent organic carbon, and (3) a few simplifying assumptions. They supported this relation by comparing 50 measurements of TOC by lab analysis to their estimates of TOC. These estimates were

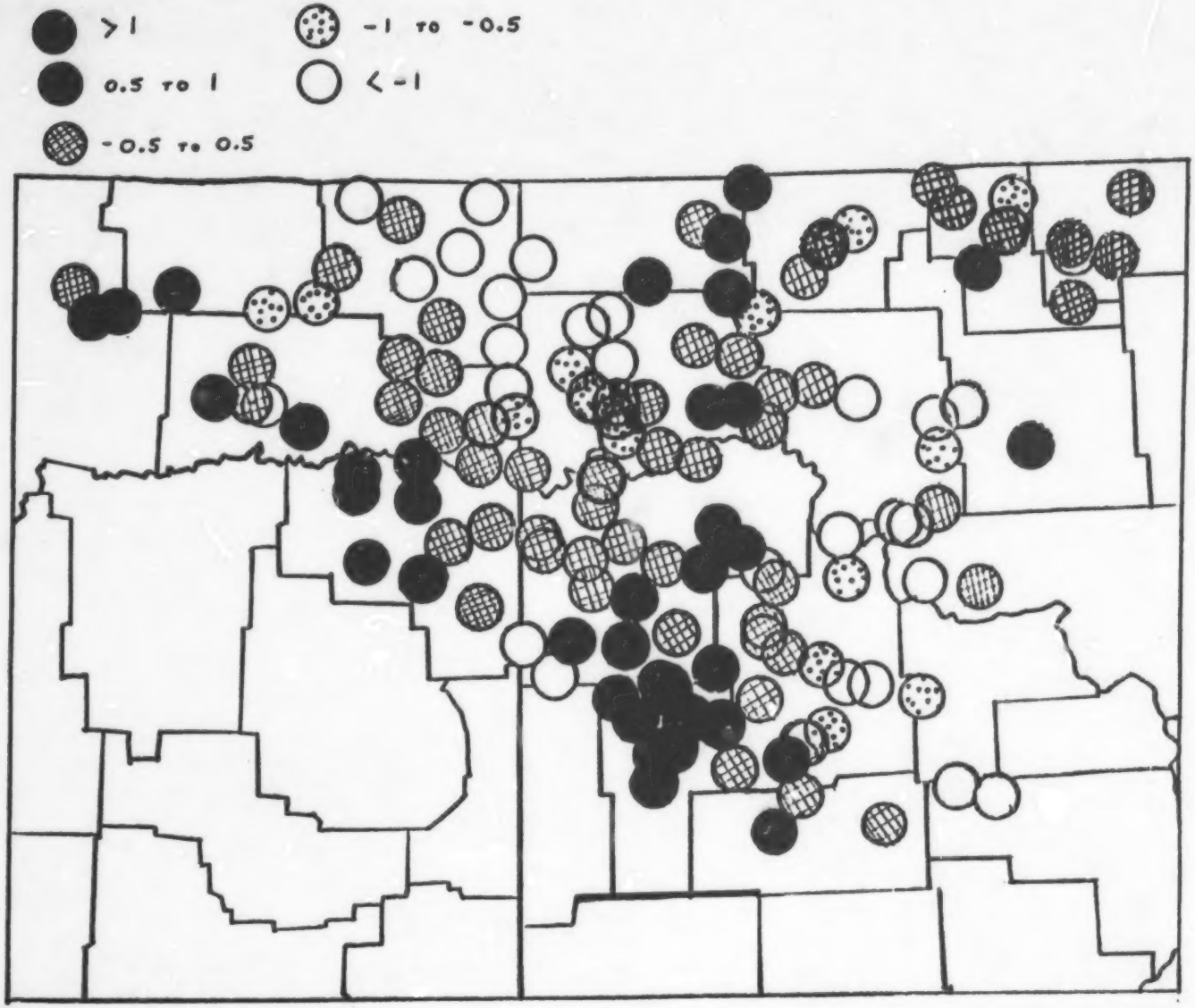


FIGURE 3

generated using their equation and density-log measurements taken from the same well or the nearest well for which data were available.

In order to further examine the relationship between TOC and density, simple linear models of the form TOC = A + B * DENSITY were fit to the data and to various subsets. Most organic-carbon analyses sampled one entire member. In some cases only part of a member was analyzed, and thus to retain comparability, an average log reading for that part of the member was used. In a few other cases the organic-carbon analysis was for both upper and lower members together, and in that case a weighted average density reading of both members was used. Some changes were made to Schmoker and Hester's data. Specifically, in two cases, a single density had been compared to two separate TOC analyses for the same interval. The less reliable TOC analysis was dropped in each of these cases. Data from 13 additional wells were added, for a total of 61 comparisons.

For density values ranging from 2.0 to 2.4 grams per cubic centimeter, a linear model relating TOC directly to density seems to be adequate (table 2). In the linear fit, the coefficient of density is negative. Thus when density values are high, TOC is low and vice versa. The relationship between TOC and density is much stronger for the upper member than for the lower member. This may well be due, at least in part, to the smaller range in values of data from the lower member. This smaller range in density and TOC values is, in turn, probably due to the more restricted horizontal and vertical range of the lower member.

Table 2.--Correlation coefficients for regressions using total organic content and density.

Model	Upper member		Lower member		Both members	
	Number of data points	R^2	Number of data points	R^2	Number of data points	R^2
TOC versus 1/density (all data)	27	0.74	26	0.36	61	0.62
TOC versus density (all data)	27	0.75	26	0.37	61	0.62
TOC versus density (<10 miles)	22	0.84	24	0.34	53	0.67
TOC versus density (<6 miles)	19	0.80	16	0.03	42	0.64

Some of the pairs of wells used in the comparison were as much as 20 miles apart. Deletion of pairs with distances greater than or equal to 10 miles strengthened the relationship for the upper member (table 2). Deletion of pairs with distances of 6 or more miles slightly weakened this relationship in the upper member. For the lower member, the relationship was slightly weakened by dropping the pairs with distances of 10 or more miles, and considerably weakened by dropping those with distances of 6 or more miles.

The weakening of the relationships when pairs with large distances were deleted can be explained partially by reference to figure 4. Many of the wells with low density-log measurements are far from the corresponding TOC wells. This is particularly true for the lower member, where 7 of 8 points with density below 2.24 grams per cubic centimeter had 6 or more miles between the compared wells. Deletion of pairs with large distances thus would tend to restrict the densities to a more narrow range. The relationship between TOC and density would be less clear when viewed over this narrow range than when viewed over the wider range of the complete data set.

Cluster and Discriminant Analysis

Further analyses performed to define the variability in the data included cluster and discriminant analysis. Clustering is done in stages. Initially, the 2 points which have the closest values on the variables used (in whatever distance measure is being used) are merged into a group. Next, using the unmerged

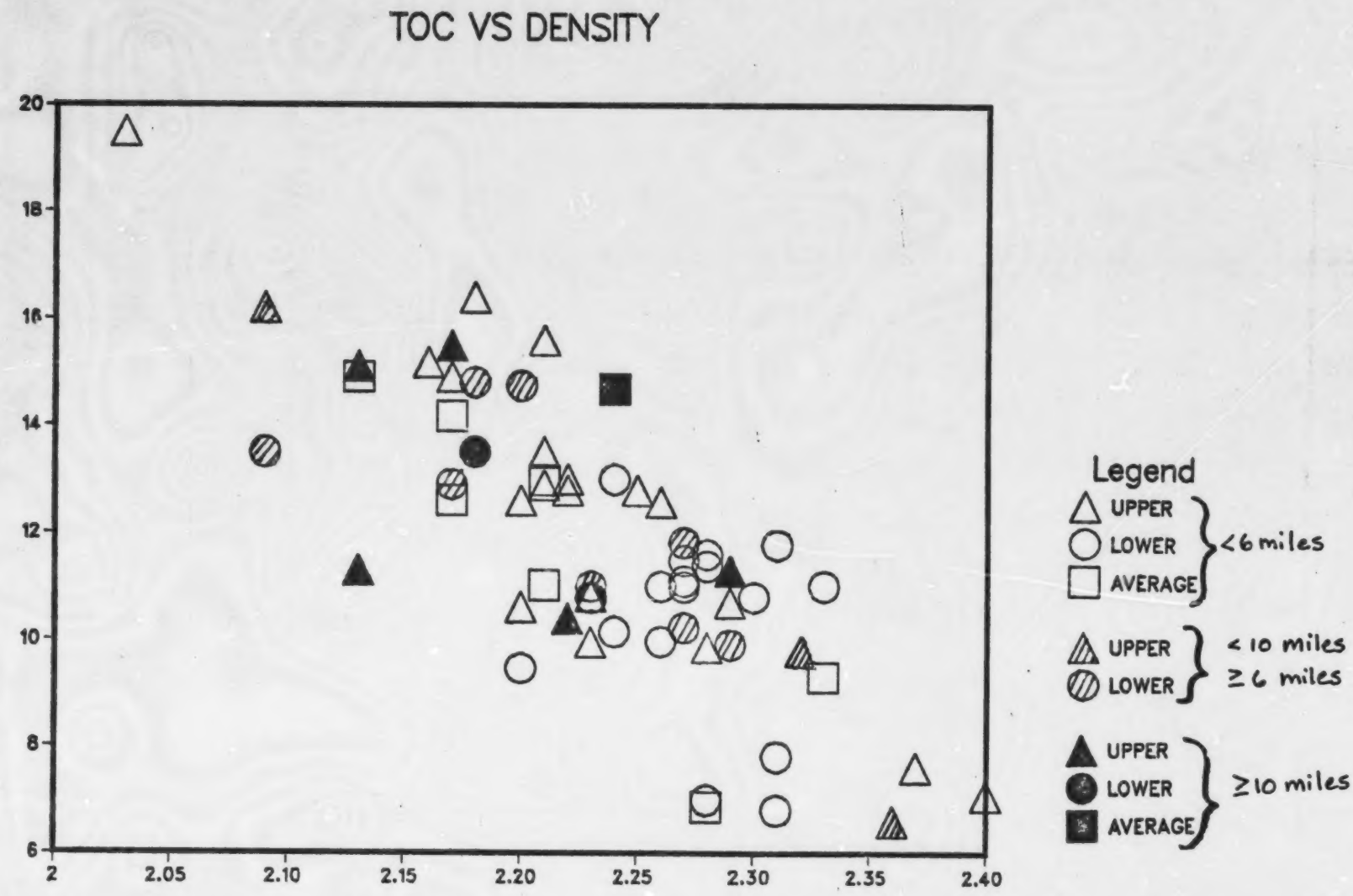

TOC VS DENSITY
20
18
16
14
12
10
8
6
2
2.05
2.10
2.15
2.20
2.25
2.30
2.35
2.40
Legend
UPPER
LOWER
AVERAGE
< 6 miles
UPPER
LOWER
< 10 miles
≥ 6 miles
UPPER
LOWER
AVERAGE
≥ 10 miles

Figure 4

points and the new group, the two closest points or point and group merge. This process continues until the points are grouped into a relatively small number of clusters. Clustering of wells was done on the upper and lower members separately. The particular clustering algorithm used was hierarchical clustering with complete linkage and Euclidean distance. Using modified density, neutron porosity, resistivity, and formation temperature, four distinct clusters were formed for each member.

Discriminant analysis was performed on the upper and lower member clusters to identify the functions which best discriminate between the four cluster groups. The discriminant functions take the form of a linear combination of the variables, the same variables used in the cluster analysis. The functions are similar for the upper and lower member. The main variables in the first discriminant function are density and neutron porosity (table 3), indicators of TOC. The main variable in the second discriminant function is formation temperature, a temperature gradient measure since depth has been removed. The plot where the value of the first discriminant function for each point is the x axis coordinate and the value of the second discriminant function is the y axis coordinate, shows that the TOC and temperature-gradient functions discriminate between groups of similar wells in the Bakken (fig. 5).

Table 3.--Discriminant function coefficients--upper member.

	First discriminant function	Second discriminant function
density	0.63	-0.30
neutron porosity	-0.65	0.15
resistivity	-0.09	0.09
temperature	0.21	0.97

Discriminant scores are the values used to classify a new point as one of the four groups. When the scores of the first discriminant function are mapped (fig. 6), they show a pattern similar to that of TOC.

Analysis of well-log data shows that variability related to depth is very strong, but that once the depth-related component is removed, present-day TOC distribution is the major control of variability. The present-day TOC distribution is due to a combination of both maturity and depositional effects which must be separated before the data can be used in a model of hydrocarbon generation.

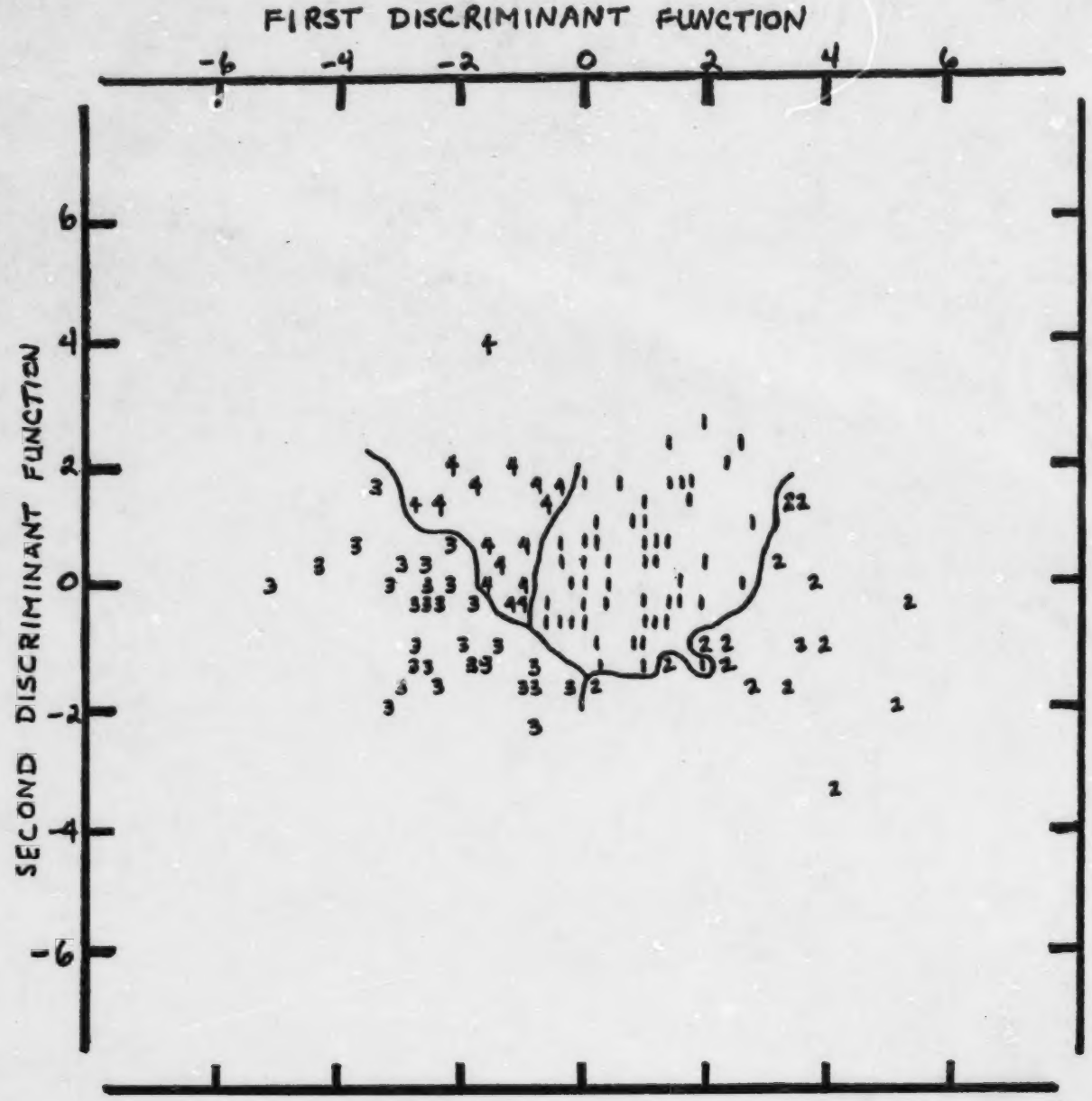

FIRST DISCRIMINANT FUNCTION
-6
-4
-2
0
2
4
6
SECOND DISCRIMINANT FUNCTION
6
4
2
0
-2
-4
-6

Figure 5

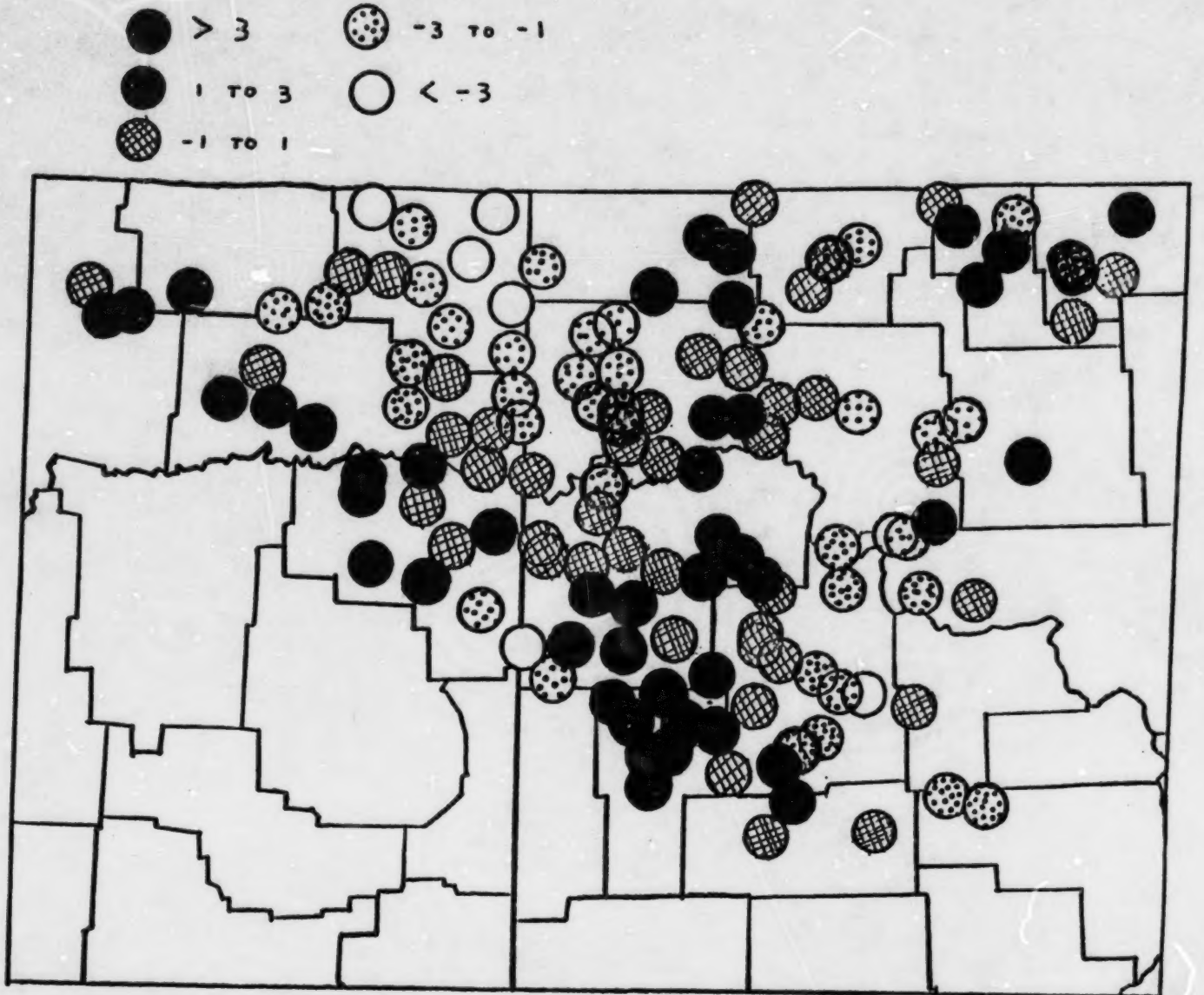


FIGURE 6

SEPARATION OF MATURITY EFFECTS FROM DEPOSITION

Regression analysis was used in this study to better separate maturity effects from original deposition. If one assumes that present-day formation temperature is a good measure of maturity, then regressing organic-carbon measurements (or some substitute such as density) on temperature should separate the maturity effects explained by the regression from original depositional effects--the residuals.

Schmoker and Hester (1983) showed that their estimated organic-carbon values decreased with formation temperature. They suggested that this was due to greater thermal maturity, and thus depletion of organic matter, with increased temperature. If TOC were negatively correlated with temperature, then density should be positively correlated with temperature. This is shown to be the case in figures 7 and 8, though the relationship is not too strong, especially for the lower member.

Residuals from this regression for the upper member show a marked areal pattern (fig. 9). The pattern is similar but weaker in the lower member. A band of positive residuals borders each member on the southwest and, less strongly, on the east. This border is broken by a patch of negative residuals in the southeast. Another patch of negative residuals is very prominent in northeast Montana in the upper member, but is considerably less prominent in the lower member.

If the changes in TOC due to maturity differences are explained by the regression, the residual pattern may reflect differences in the pre-maturation concentrations of organic material. Large positive residuals reflect areas of higher density and thus lower original TOC. Large negative residuals reflect areas of lower density and thus higher original TOC. The original pattern of organic matter concentration would thus be low organic content near the shoreline, increasing offshore to high organic content in the northeast Montana area as would be consistent with a reasonable depositional model. An area of high organic content to the southeast may suggest that the original shoreline in this area has been eroded.

An alternative explanation for this residual pattern is that a linear model is inappropriate for relating density and temperature. Because chemical reactions increase exponentially with increasing temperature, hydrocarbon generation might be expected to increase exponentially with increasing temperature (Waples, 1980). We reject this explanation for the following reason: For high or low temperatures, a linear model would be expected to overestimate the TOC, while for moderate temperatures the linear model would slightly underestimate the TOC (fig. 10). Thus, one could expect negative TOC residuals (positive density residuals) in areas of high or low temperature and positive TOC residuals (negative density residuals) in areas of moderate temperature. Comparison of the areal residual pattern (fig. 9) with formation temperature (fig. 11) shows that this is not so.

REGRESSION OF DENSITY ON FORMATION TEMPERATURE
UPPER MEMBER

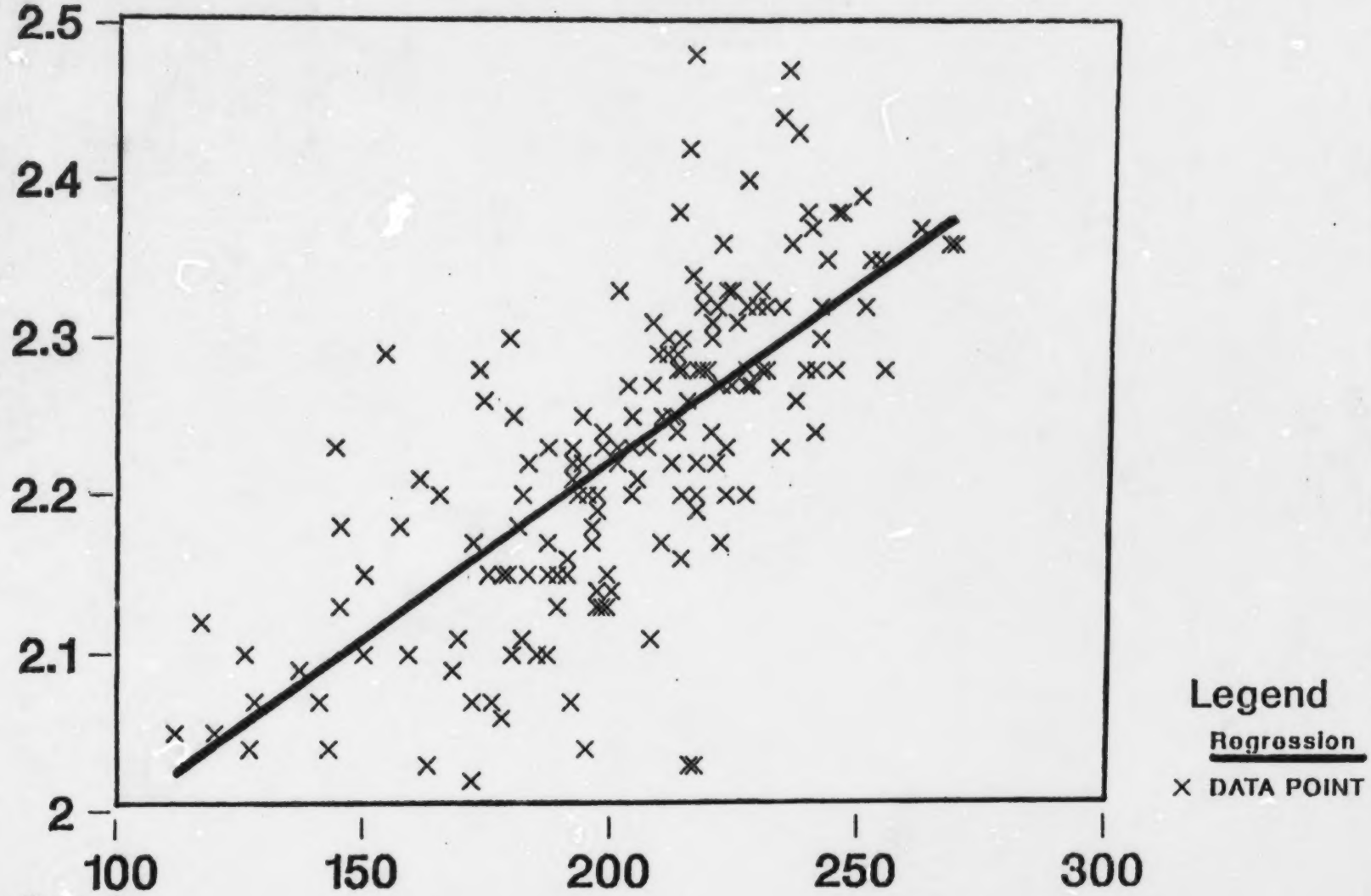


Figure 7

REGRESSION OF DENSITY ON FORMATION TEMPERATURE
LOWER MEMBER

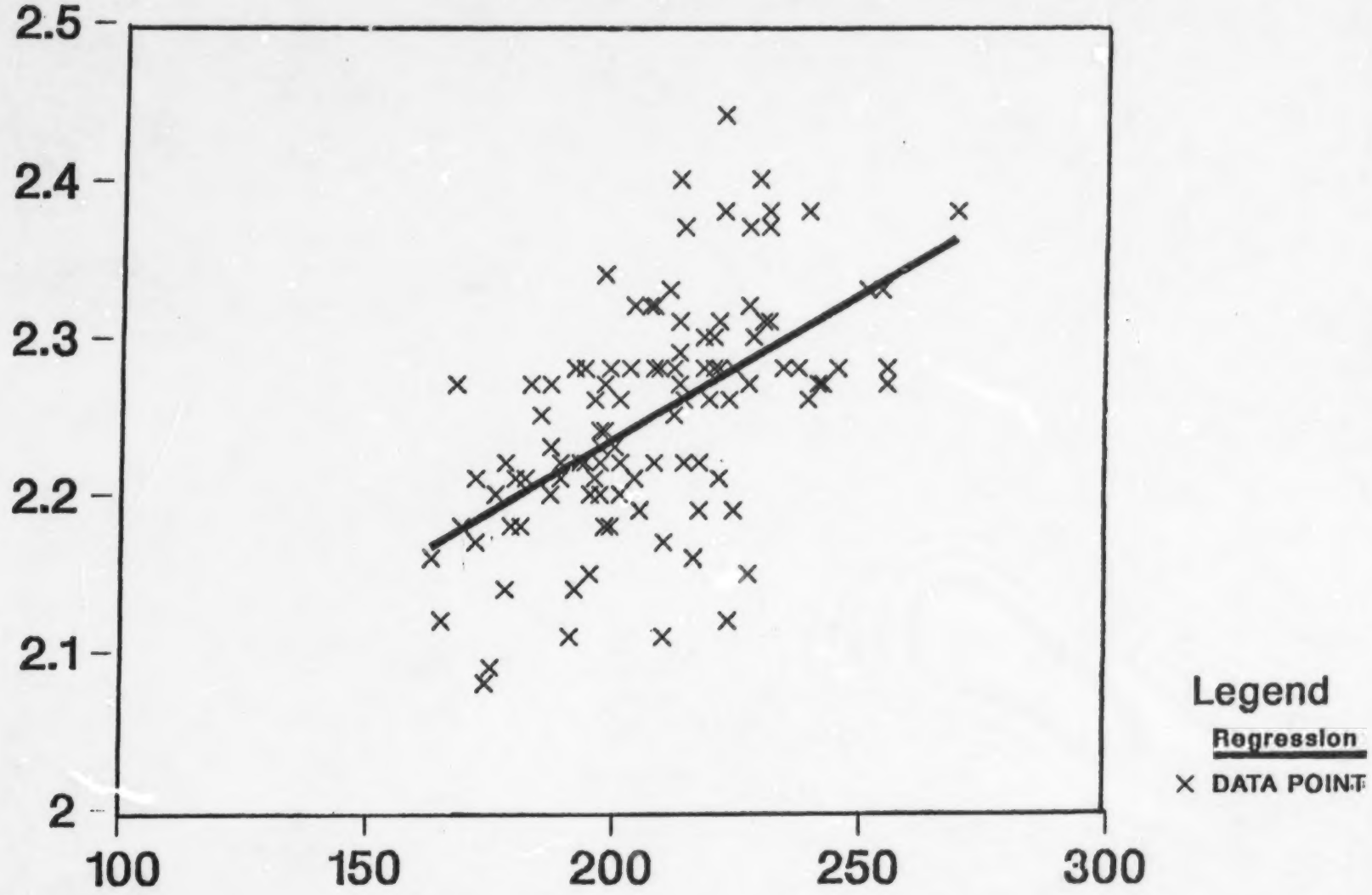


Figure 8

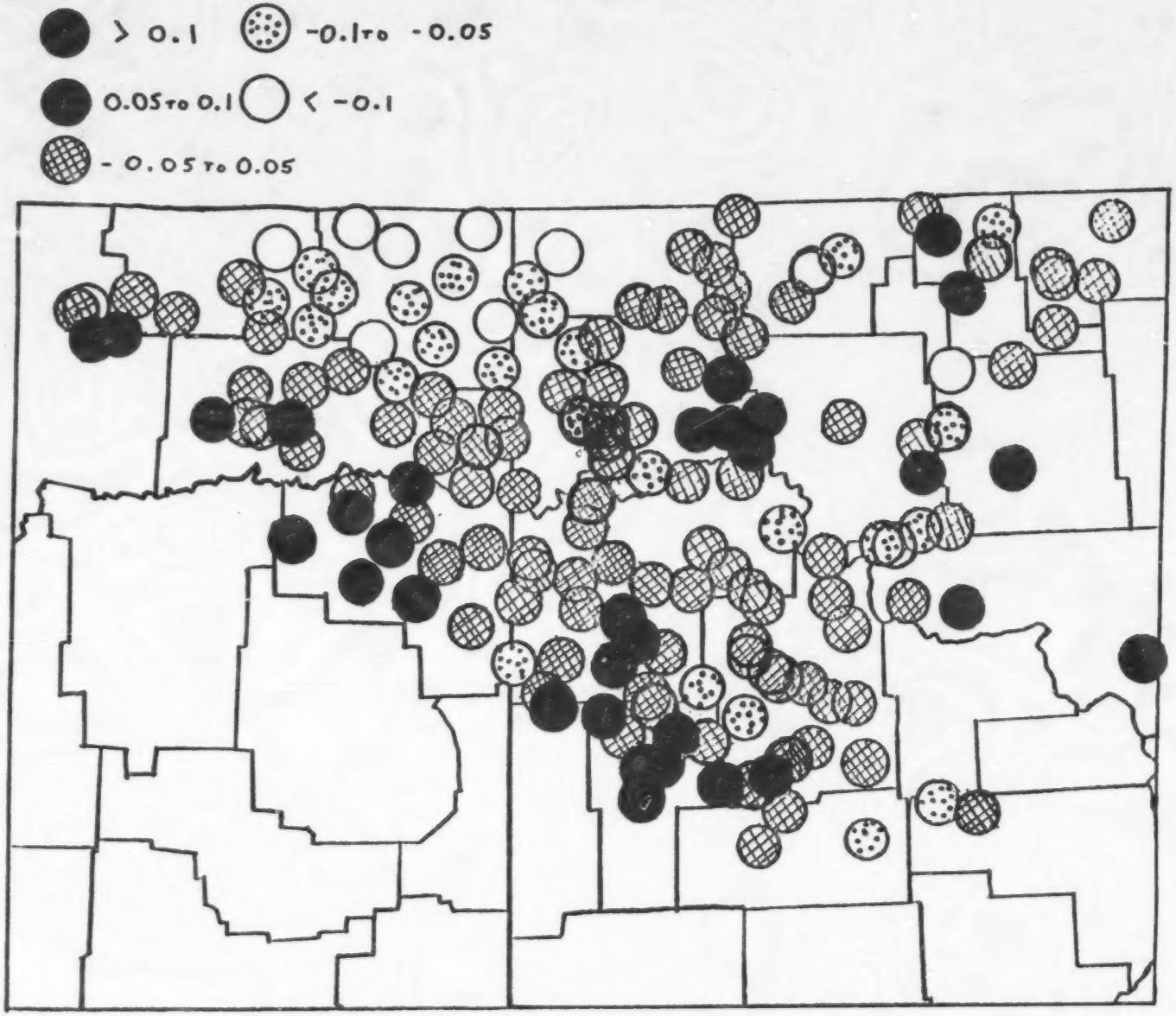


FIGURE 9

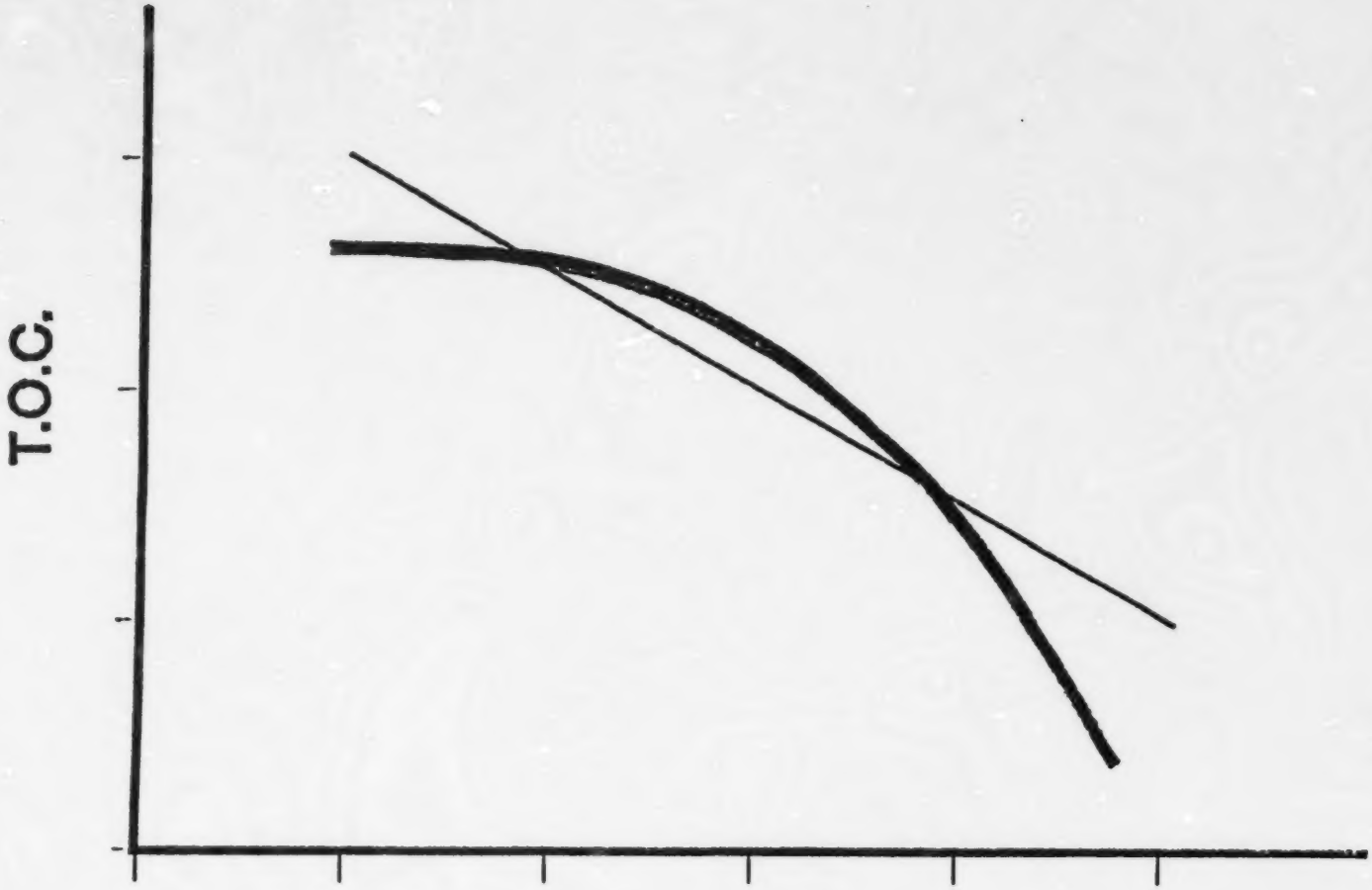


Figure 10

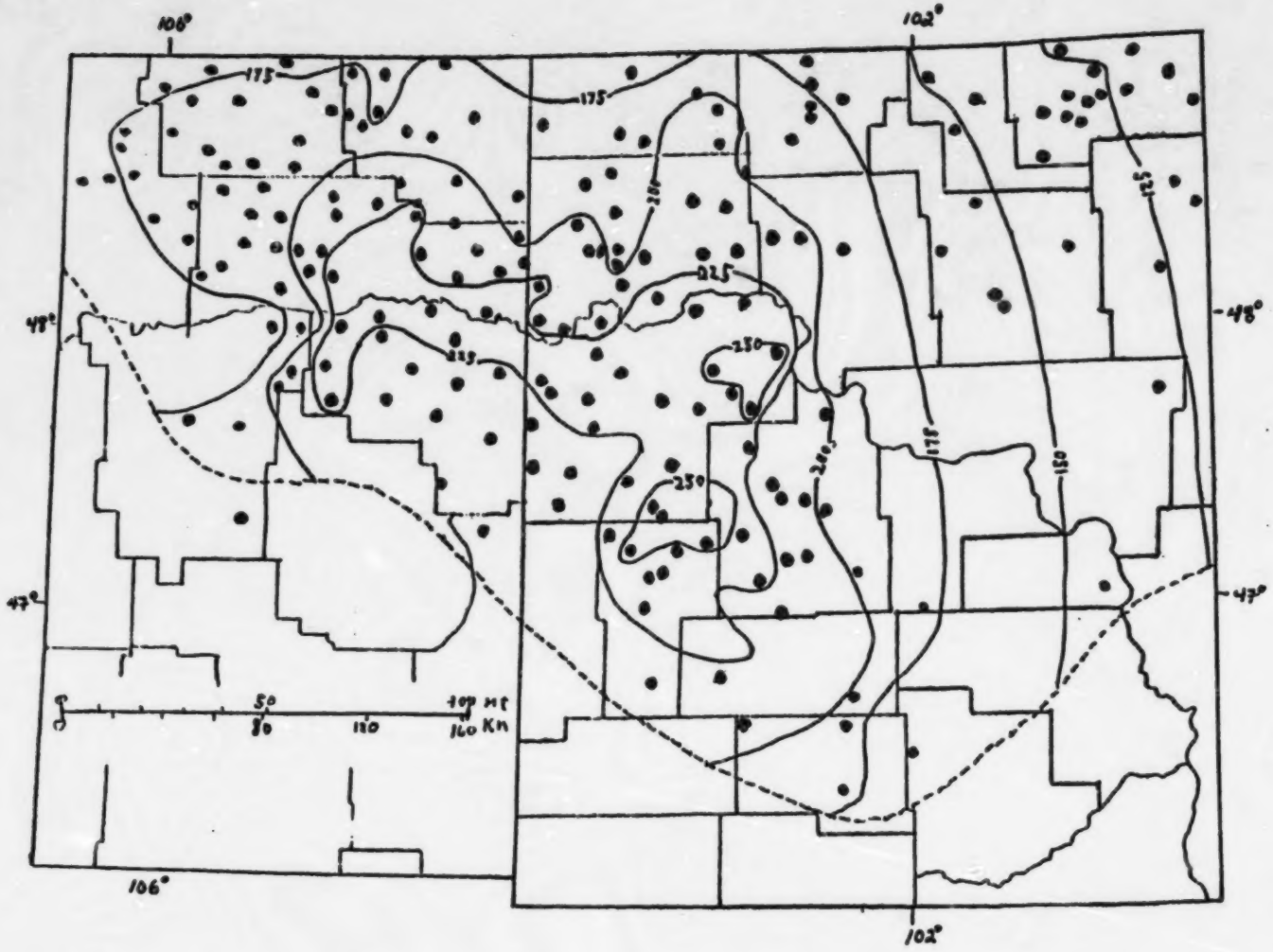


FIGURE 11

The residual pattern thus can be better explained by depositional differences than by an inappropriate model.

CONCLUSIONS

In the upper member of the Bakken Formation, the pattern of TOC before maturity indicates low values on the southwest and east edges, increasing offshore to high TOC in northeastern Montana (fig. 12). This pattern reflects the original deposition of the formation. The pattern of TOC before maturity for the lower member is similar but not as well-defined.

The current pattern of TOC in both the upper and lower members is high organic content in northeast Montana and along the east edge of the basin (fig. 13). Low organic content occurs along the southwest edge of the basin and along the Nesson anticline.

The contrast between these two patterns reflects the effect of maturity. The part of the basin along the anticline has been subjected to higher temperatures. This explains the pre- to post-maturation change from moderate to low organic content as compared with other values in the formation. The change from low to high relative organic content along the eastern edge of the basin can be explained by much lower temperatures in that area.

Separation of maturity effects from the original deposition is achieved using regression analyses on the well-log data. Studies similar to this may yield equally valuable results on marine shales in other basins.

ACKNOWLEDGMENTS

Much of the data used in this study was collected and edited by J. W. Schmoker and T. C. Hester. Thanks go to them for providing the data and for the many helpful comments during this study. We also thank L. C. Price for making his geochemical data available to us, and the reviewers, J. C. Houghton and J. W. Schmoker, for their suggestions.

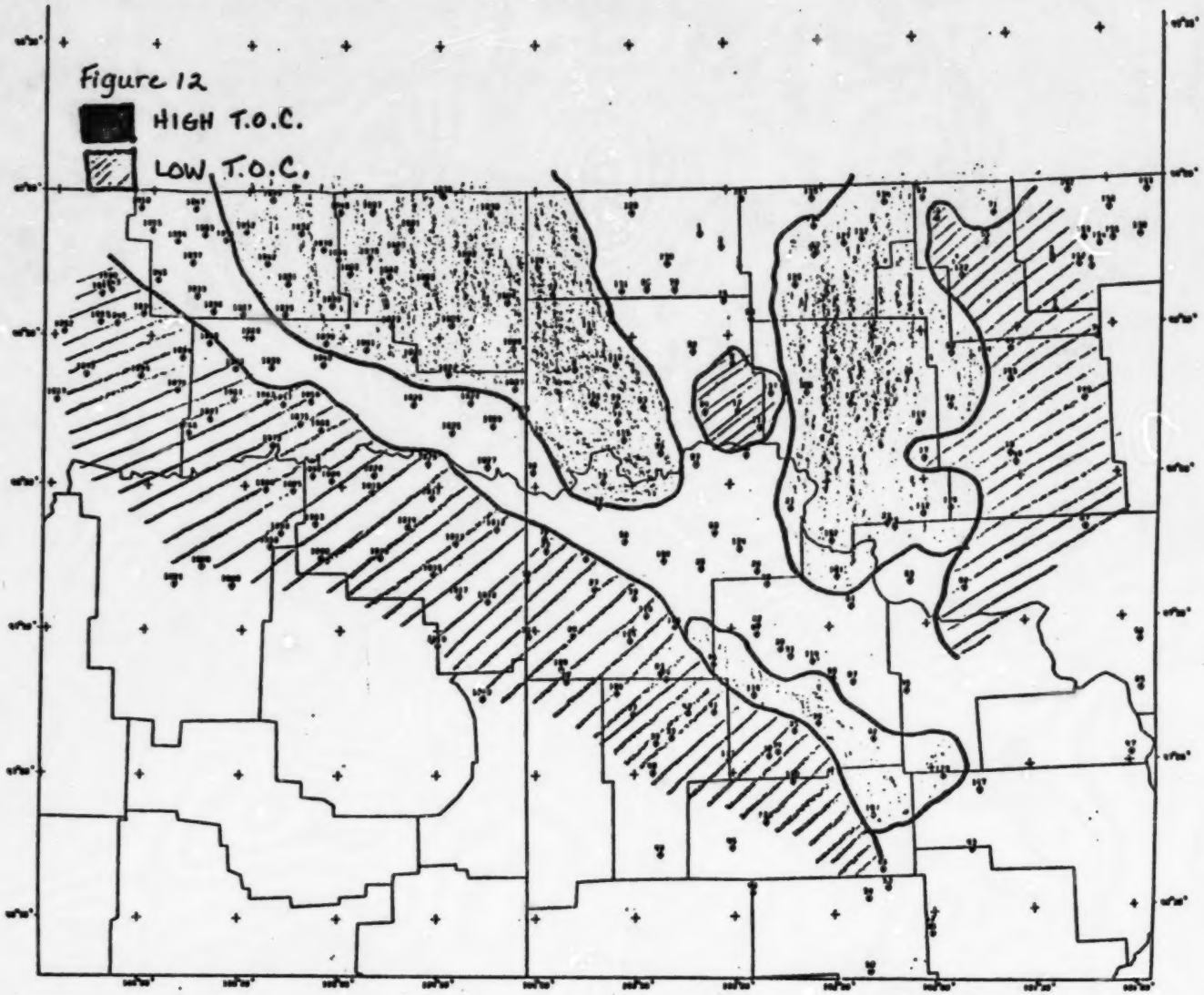


Figure 12
HIGH T.O.C.
LOW T.O.C.

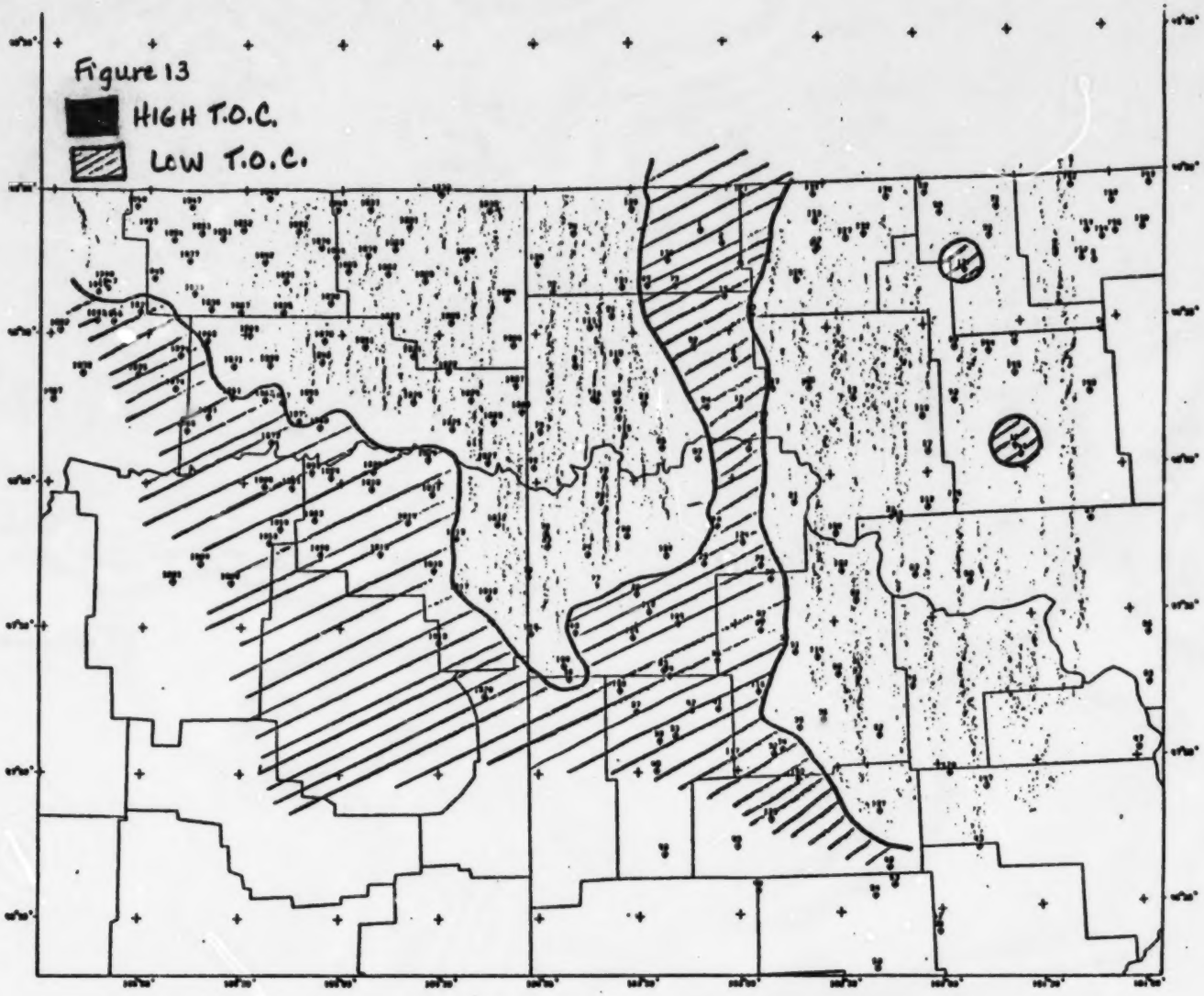


Figure 13
HIGH T.O.C.
LOW T.O.C.

REFERENCES

Dow, W. G., 1974, Application of oil-correlation and source-rock data to exploration in Williston basin: American Association of Petroleum Geologists Bulletin, v. 58, no. 7, p. 1253-1262.

Green, P. E., 1978, Analyzing Multivariate Data: Hinsdale, Illinois, The Dryden Press, 519 p.

Hester, T. C., and Schmoker, J. W., 1983, Preliminary log-derived maps of the Bakken Formation, North Dakota and Montana portions of Williston basin: U.S. Geological Survey Open-File Report 83-717, 25 p.

Meissner, F. F., 1978, Petroleum geology of the Bakken Formation, Williston basin, North Dakota and Montana, in The economic geology of the Williston basin: Montana Geological Society 24th Annual Conference; 1978 Williston Basin Symposium, p. 207-227.

Price, L. C., Ging, T., Daws, T., Love, A., Pawlewicz, M., and Anders, D., 1984, Organic metamorphism in the Mississippian-Devonian Bakken Shale, North Dakota portion of the Williston basin: Rocky Mountain Association of Geologists Symposium Volume on Sources for Oil and Gas in the Rocky Mountain Area,(in press).

Schmoker, J. W., and Hester, T. C., 1983, Organic carbon in Bakken Formation, United States portion of Williston basin: American Association of Petroleum Geologists Bulletin, v. 67, no. 12, p. 2165-2174.

Waples, D. W., 1980, Time and temperature in petroleum formation application of Lopatin's method to petroleum exploration: American Association of Petroleum Geologists Bulletin, v. 64, no. 6, p. 916-926.

Webster, R. L., 1982, Analysis of petroleum source rocks of the Bakken Formation (Devonian and Mississippian) in North Dakota: University of North Dakota, unpublished Masters thesis, 150 p.

Williams, J. A., 1974, Characterization of oil types in Williston basin: American Association of Petroleum Geologists Bulletin, v. 58, no. 7, p. 1243-1252.